AFRL-IF-RS-TR-2005-126
Final Technical Report
April 2005

STOCHASTIC FLUCTUATIONS IN GENE REGULATION

Boston University

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-126 has been reviewed and is approved for publication

APPROVED: /s/
PETER J. COSTIANES
Project Engineer

FOR THE DIRECTOR: /s/
JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
April 2005

3. REPORT TYPE AND DATES COVERED
Final **Aug 01 – Aug 04**

4. TITLE AND SUBTITLE

STOCHASTIC FLUCTUATIONS IN GENE REGULATION

5. FUNDING NUMBERS
G - F30602-01-2-0579
PE - N/A
PR - BIOC
TA - M2
WU - 95

6. AUTHOR(S)
J. Collins
T. Elston
A. van Oudenaarden

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Primary: Boston University, Boston MA 02215

Sub: University of North Carolina, Chapel Hill NC 27599
Massachusetts Institute of Technology, Cambridge MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFED
525 Brooks Road
Rome NY 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2005-126

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Peter J. Costianes/IFED/(315) 330-4030 Peter.Costianes@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

This project developed improved theoretical and simulation techniques that take account of stochastic fluctuations in gene expression and lead to reliable predictions of complex cell behaviors. These techniques and associated models were tested and refined through experimental studies using engineered gene circuits. Through this integrated approach, a hierarchical simulation approach to describe, predict and control complex gene networks in living cells are developed and incorporated into BioSPICE.

14. SUBJECT TERMS
Simulations, software, gene expression, stochasticity, gene networks

15. NUMBER OF PAGES 22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Contents

1 Project Summary

Although ignored by most molecular biologists who commonly characterize gene expression levels, the reality is that the level of expression of the same gene can vary enormously from one cell to another within a genetically-identical cell population. It has been shown that these fluctuations are not simply a by-product of the regulatory process and that they can contribute significantly to the control of cell function. For example, some organisms utilize fluctuations to introduce diversity into a population, as occurs during the lysis-lysogeny transition in λ phage, while in others stability against fluctuations is essential in gene regulatory cascades controlling processes such as differentiation. This project developed improved theoretical and simulation techniques that lead to reliable predictions of complex cell behaviors. Moreover, this project developed quantitative models of gene regulation that correctly incorporate stochastic fluctuations that are inherent in all gene networks and currently limit the usefulness of existing modeling approaches. Computational models were tested and refined through experimental studies using engineered gene circuits. Through this integrated approach, a hierarchical simulation approach was developed and incorporated into BioSPICE in order to enhance the user's ability to describe, predict and control complex gene networks in living cells.

2 Significant Accomplishments

The significant accomplishments for the project are summarized below:

- GRASS v2.0, Gene Regulatory Anaylsis and Stochastic Simulation, was developed, successfully integrated and released as part of BioSPICE v2.0.
- SDEsolver, a general purpose biochemical reaction simulator, was developed, successfully integrated and released as part of BioSPICE v2.0.
- BioNetS, a biochemical network simulator, was developed, successfully integrated and released as part of BioSPICE v3.0. BioNetS handles discrete, continuous, and mixed models. It also provides a semi-implicit method for stiff systems, and is optimized for speed and validated for accuracy.
- Models for sources of noise (transcription and translation) in gene expression systems (single-gene systems and cascade networks of genes) were developed and experimentally validated using engineered gene networks.
- Models of synthetic engineered promoters were developed and experimentally validated using engineered gene circuits.
- A synthetic engineered promoter, which could be independently activated and repressed by two different proteins, was constructed and used to validate the developed models.
- Gene expression systems (single-gene systems and cascade networks of genes) were constructed and used to validate models for sources of noise (transscription and translation) in gene expression.
- A "noise generator" that can be used to produce gene regulatory signals with variable signal-to-noise ratios was developed.
- Prototype programmable cells in which engineered regulatory modules were integrated with the cell's natural circuitry were developed.

3 Software for Stochastic Modeling of Biochemical Networks

3.1 Background

Mathematical modeling of complex biological networks has a lengthy history. In the past, the standard approach for modeling these systems has been to derive ordinary differential equations (ODEs) based on the law of mass action for the concentrations of the biochemical species involved in the network. Experimental studies have demonstrated, however, that stochastic effects can be significant in cellular reactions, particularly in the case of transcriptional regulation, where generally there are two copies of each gene and the number of messenger RNA (mRNA) molecules can be small. A number of recent experimental and modeling studies have addressed the role of fluctuations in gene expression. Many modeling studies have employed the well-established Gillespie Monte Carlo algorithm or one of its more recent variants. These algorithms offer an exact solution to the stochastic evolution of chemical systems, but they are computationally very expensive. A much more efficient approach is to approximate the species as continuous variables and formulate the problem in terms of stochastic differential equations (SDEs), often referred to as chemical Langevin equations. This approximation works remarkably well for many cases, even when the number of particles involved is as small as 10, and the resulting simulations can run orders of magnitude more quickly than the discrete Monte Carlo approach. In other cases, when some or all of the particle numbers are very small, the system may need to be modeled using the discrete approach, or a hybrid method in which some species are treated discretely while others are evolved using the continuum approximation. With the increasing interest in formulating accurate models of large biochemical networks, there is a need for reliable software packages that correctly incorporate stochastic effects, yet are fast enough to simulate large interconnected sets of reacting species (as found, for example, in signaling cascades or genetic regulatory networks). The BIOchemical NETwork Stochastic Simulator, "BioNetS," was developed to meet this need. BioNetS is capable of performing full discrete simulations using an efficient implementation of the Gillespie algorithm. It is also able to set up and solve the chemical Langevin equations, which are a good approximation to the discrete dynamics in the limit of large abundances. Finally, BioNetS can handle hybrid models in which chemical species that are present in low abundances are treated discretely, whereas those present at high abundances are handled continuously. Thus, the user can pick the simulation method that is best suited to their needs. All aspects of the software are highly optimized for efficiency.

In the Implementation section, the mathematical background for the Gillespie method, chemical Langevin equations and hybrid models is presented, along with a discussion of the numerical algorithms used in BioNetS. Under Results and Discussion, a brief introduction to BioNetS is provided along with several examples. The examples serve two purposes: (1) to illustrate how to use the software, and (2) to verify its efficiency and accuracy.

3.2 Implementation

The mathematical methodology on which BioNetS is built is developed first.

3.2.1 Discrete Reactions and the Gillespie Algorithm

BioNetS makes use of elementary reactions (zeroth, first and second order). The following examples illustrates each type of reaction:

$$\emptyset \underset{\delta}{\overset{\gamma}{\rightleftharpoons}} \mathcal{A} \tag{1}$$

$$\mathcal{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{B} \tag{2}$$

$$\mathcal{A} + \mathcal{B} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \mathcal{A}\_\mathcal{B} \tag{3}$$

$$\mathcal{V} \underset{k_6}{\overset{k_5}{\rightleftharpoons}} \mathcal{V} + \mathcal{V} \tag{4}$$

In the above reactions, the calligraphic letters denote a single molecule of a chemical species. The number of molecules of a particular species in the system at time t is denoted with uppercase letters (e.g., $A(t)$, $B(t)$, $A\_B(t)$, and $V(t)$). All the rate constants, γ, δ, and k_1-k_6, have units of per time. Eq. 1 represents a process in which a molecule $\mathcal{A}$ is produced when the reaction proceeds in the forward direction and is degraded in the reverse direction. In the forward direction, the reaction is zeroth order and proceeds with an average rate of γ. In the backward direction, the reaction is first order, and the average rate of degradation is $\delta A(t)$. The forward reaction in Eq. 2 represents a process in which chemical species $\mathcal{A}$ is converted to species $\mathcal{B}$. In this case, $\mathcal{A}$ and $\mathcal{B}$ might represent two different conformations of the same molecule. In Eq. 2 both the forward and backward reactions are first order because the reaction rates are proportional to the respective concentrations. The forward reaction given in Eq. 3 is a second-order reaction in which an $\mathcal{A}$ molecule and a $\mathcal{B}$ molecule come together to form the complex $\mathcal{A}\_\mathcal{B}$. The average rate for the reaction is $k_1A(t)B(t)$. The backward reaction is a first-order reaction in which $\mathcal{A}\_\mathcal{B}$ dissociates at an average rate of $k_2A\_B(t)$. In Eq. 4 the forward reaction produces a molecule $\mathcal{V}$. The difference between this reaction and the forward reaction in Eq. 1 is that the average rate is $k_3V(t)$. This leads to exponential growth of $V(t)$. This reaction is particularly useful if $V(t)$ is interpreted as the cell volume. In the backward reaction, two $\mathcal{V}$ molecules come together and degrade one of the $\mathcal{V}$ molecules. The average rate for this reaction is $k_4V(t)(V(t)-1)$. The $V(t)-1$ term arises because two of $V(t)$ molecules must be chosen to react. This type of term also arises in reactions that produce homodimers. This reaction eventually stops the exponential growth of V. The net effect of these two reactions is to produce logistic growth. The total average reaction rate for the set of reactions given in Eqs. 1-4 is

$$\begin{aligned} \mu(t) &= \gamma + \delta A(t) + k_1A(t) + k_2B(t) + \\ &\quad k_3A(t)B(t) + k_4A\_B(t) + k_5V(t) + k_6V(t)(V(t)-1) \\ &= \sum_{i=1}^{4}(F_i + B_i) \end{aligned} \tag{5}$$

where F_i and B_i are the average forward and backward rates, respectively, for the ith reaction.

For the rest of this section, it is assumed that the volume of the cell is not changing and only Eqs. 1-3 are considered. In the examples, a case in which the volume is changing is considered. If $A(t)$, $B(t)$ and $A\_B(t)$ are present in large numbers, then the law of mass action can be applied to derive equations that govern the concentrations $[A] = A(t)/V$, $[B] = B(t)/V$ and $[AB] = A\_B(t)/V$, where V is the cell volume. These equations are

$$\frac{d[A]}{dt} = -(k_3'[B] + k_1 + \delta)[A] + k_4[AB] + k_2[B] + \gamma' \tag{6}$$

$$\frac{d[B]}{dt} = -(k_3'[A] + k_2)[B] + k_1[A] + k_4[AB] \tag{7}$$

$$\frac{d[AB]}{dt} = k_3'[A][B] - k_4[AB] \tag{8}$$

The primed rate constants indicate that they have been appropriately scaled by the volume (i.e, $k_3' = k_3V$ and $\gamma' = \gamma/V$), and, therefore, have units of either per time per concentration or concentration per time. Note that to convert to units of molar, one also has to appropriately scale the rate constants by Avagadro's number. Eqs. 6 - 8 represent a macroscopic description of the process, because they ignore fluctuations in the concentration that arise from the stochastic nature of chemical reactions.

In general, $A(t)$, $B(t)$ and $A\_B(t)$ are random variables that take on any nonnegative integer value. The Gillespie algorithm can be used to generate sample paths of the process. This algorithm assumes that the random time ΔT_i, between the ith and $i+1$ reaction, is exponentially distributed. For the simple example given by Eqs. 1 - 3, the mean waiting time between reactions, which characterizes the exponential distribution, is $\mu_{\Delta T_i} = \gamma + \delta A(t_i) + k_1A(t_i) + k_2B(t_i) + k_3A(t_i)B(t_i) + k_4A\_B(t_i)$, where t_i is the time at which the ith reaction occurred. Therefore, $t_{i+1} = t_i + \Delta T_i$.

Once the time at which the next reaction occurrs has been determined, the following probabilities are used to determine which reaction occurred:

$$\begin{aligned}
\Pr[\emptyset \xrightarrow{\gamma} \mathcal{A}] &= \frac{\gamma}{\mu_{\Delta T_i}} && (9)\\
\Pr[\mathcal{A} \xrightarrow{\delta} \emptyset] &= \frac{\delta A(t_i)}{\mu_{\Delta T_i}} && (10)\\
\Pr[\mathcal{A} \xrightarrow{k_1} \mathcal{B}] &= \frac{k_1 A(t_i)}{\mu_{\Delta T_i}} && (11)\\
\Pr[\mathcal{B} \xrightarrow{k_2} \mathcal{A}] &= \frac{k_2 B(t_i)}{\mu_{\Delta T_i}} && (12)\\
\Pr[\mathcal{A} + \mathcal{B} \xrightarrow{k_3} \mathcal{A\_B}] &= \frac{k_3 A(t_i) B(t_i)}{\mu_{\Delta T_i}} && (13)\\
\Pr[\mathcal{A\_B} \xrightarrow{k_4} \mathcal{A} + \mathcal{B}] &= \frac{k_4 A\_B(t_i)}{\mu_{\Delta T_i}} && (14)
\end{aligned}$$

Once the reaction has been determined, the chemical species are updated accordingly. As discussed in the Numerical Methods section, BioNetS uses an efficient implementation of the Gillespie algorithm.

Another description of discrete stochastic processes is achieved through use of the master equation that governs how the probabilities of the various random variables in the process evolve in time. Let $p_{a,b,ab}(t) = \Pr[A(t) = a, B(t) = b, A\_B(t) = a\_b]$, then $p_{a,b,a\_b}(t)$ satisfies the master equation

$$\begin{aligned}
\frac{dp_{a,b,a\_b}}{dt} &= -[\gamma + (\delta + k_1)a + k_2 b + k_3 ab + k_4 a\_b] p_{a,b,a\_b} + \gamma p_{a-1,b,a\_b} + \delta(a+1) p_{a+1,b,a\_b} \\
&\quad + k_1(a+1) p_{a+1,b-1,a\_b} + k_2(b+1) p_{a-1,b+1,a\_b} \\
&\quad + k_3(a+1)(b+1) p_{a+1,b+1,a\_b-1} + k_4(a\_b+1) p_{a-1,b-1,a\_b+1} \qquad (15)
\end{aligned}$$

The master equation is the starting point for deriving various approximate schemes for describing the system. In the next section, an approximate scheme that is valid in the limit of large, but finite molecule numbers, is discussed. The simplest approximation scheme is achieved by considering the first moments of the process. Over bars will be used to denote averaging. For example, $\bar{A}(t) = \sum_{a,b,a\_b} a p_{a,b,a\_b}(t)$. Eq. 15 can be used to derive equations that govern the time evolution of all the first moments. Because of the second-order reaction in Eq. 3, the equations for the means are coupled to the second moments. In fact, the nth moment equations contain terms that involve the $n+1$ moments. Thus, there is no closure to the system. The simplest closure scheme is to assume that all moments factorize (e.g., $\overline{A^2} = \overline{A}^2$). This represents the macroscopic limit in which fluctuations are ignored. In this limit, Eqs. 6-8 are recovered from the master equation.

The Diffusion Limit and the Chemical Langevin Equations

The general form of the master equation for a system consisting of N chemical species and M reactions is

$$\frac{dp_{\mathbf{n}}}{dt} = \sum_{i=1}^{M}(F_i + B_i)p_{\mathbf{n}} + \sum_{i=1}^{M} F_i p_{\mathbf{n}-\delta_{\mathbf{i}}} + \sum_{i=1}^{M} B_i p_{\mathbf{n}+\delta_{\mathbf{i}}} \tag{16}$$

where $\mathbf{n}$ is a N-dimensional vector of species numbers, F_i and B_i are the backward and forward rates for the ith reaction, and the vectors δ_i contain the stoichiometric constants for the ith reaction. For the simple model given by Eqs. 1 - 3, $N = 3$, $M = 3$, and $p_{\mathbf{n}}(t) = \Pr[A(t) = n_1, B(t) = n_2, \text{and} A\_B(t) = n_3]$. The forward and backward rates are $F_1 = \gamma$, $B_1 = \delta n_1$, $F_2 = k_1 n_1$, $B_2 = k_2 n_2$, $F_3 = k_3 n_1 n_2$, and $B_3 = k_4 n_3$. and the δ_i vectors are the rows of the stoichiometric matrix

$$\mathbf{\Delta} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \tag{17}$$

The (i, j) element in the above matrix represents the change in the jth chemical species when the ith reaction proceeds in the forward direction.

If the molecule numbers are large as compared to 1, then the master equation Eq. 16 can be approximated by the continuous process

$$\frac{\partial \rho(\mathbf{n}, t)}{\partial t} = -\sum_{j}^{N} \frac{\partial}{\partial n_j} A_j(\mathbf{n})\rho(\mathbf{n}, t) + \frac{1}{2}\sum_{j,k}^{N} \frac{\partial^2}{\partial_{n_j}\partial_{n_k}} D_{j,k}(\mathbf{n})\rho(\mathbf{n}, t) \tag{18}$$

where

$$A_j(\mathbf{n}) = \sum_{i}^{M} \Delta_{j,i}(F_i - B_i) \tag{19}$$

$$D_{j,k}(\mathbf{n}) = \sum_{i}^{M} \Delta_{i,j}\Delta_{i,k}(F_i + B_i) \tag{20}$$

This result can be derived in several ways. One method is to note that Eq. 15 represents a second-order finite differencing of Eq. 18, with a grid size of 1. Another method is to make use of the shift operator

$$f(n + k) = \exp(k\frac{\partial}{\partial n})f(n) = \sum_{j=0}^{\infty} \frac{k^j}{j!}\frac{\partial^j}{\partial n^j} f(n) \tag{21}$$

where $f(n)$ is an arbitrary smooth function and for our purposes k is an integer. If the shift operator is used in Eq. 15, the diffusion limit is achieved when the Taylor series expansion given in Eq. 21 is truncated at $j = 2$.

Sample paths consistent with Eq. 18 can be generated using the following set of SDEs

$$\frac{dN_j}{dt} = A_j(\mathbf{N}) + \sum_{k=1}^{M} \Delta_{k,j} \sqrt{F_k(\mathbf{N}) + B_k(\mathbf{N})} w_k(t) \tag{22}$$

where the $w_k(t)$ are independent Gaussian white noise processes. These equations are often referred to as the chemical Langevin equations. For Eqs. 1 - 3, the explicit form of the SDEs are

$$\begin{aligned} \frac{dA}{dt} &= -(k_1 + k_3 B + \delta)A + k_2 B + k_4 A\_B + \gamma' + \sqrt{\delta A + \gamma} w_1(t) \\ &\quad - \sqrt{k_1 A + k_2 B} w_2(t) - \sqrt{k_3 AB + k_4 A\_B} w_3(t) \qquad (23) \\ \frac{dB}{dt} &= -(k_2 + k_3 A)B + k_1 A + k_4 A\_B + \sqrt{k_1 A + k_2 B} w_2(t) \\ &\quad - \sqrt{k_3 AB + k_4 A\_B} w_3(t) \qquad (24) \\ \frac{dA\_B}{dt} &= -k_4 A\_B + k_3 AB + \sqrt{k_3 AB + k_4 A\_B} w_2(t) \qquad (25) \end{aligned}$$

BioNetS generates numerical solutions to the SDEs given by Eq. 22 using either an explicit or semi-implicit Euler method. The form of these methods is

$$\begin{aligned} N_j(t + \Delta t) &= N_j(t) + \Delta t A_j(\mathbf{N}(t + \epsilon \Delta t)) \\ &\quad + \sqrt{\Delta t} \sum_{k=1}^{M} \Delta_{k,j} \sqrt{F_k(\mathbf{N}(t)) + B_k(\mathbf{N}(t))} Z_k(t) \qquad (26) \end{aligned}$$

where $\epsilon = 0$ for the explicit method and $\epsilon = 1$ for the semi-implicit method and the $Z_k(t)$ are independent standard normal random variables. The advantage of using the chemical Langevin equations is that in the appropriate parameter regime, numerical solutions to the set of SDEs given by Eq. 22 can be generated much more efficiently than using the Gillespie algorithm. This point is expanded upon in the Results section. Higher-order numerical algorithms for SDEs are available, but the noise structure of the chemical Langevin equations makes these schemes very cumbersome to implement. In the Results section, the Euler methods given by Eq. 26 are verified to be sufficient to produce reliable results. Note that the $\mathbf{\Delta}$ matrix is generally sparse, and BioNetS takes advantage of this sparseness to optimize the efficiency of the two Euler methods (see Numerical Methods, below).

3.2.2 Hybrid Schemes

It is often desirable to allow some of the chemical species to be treated as continuous random variables and some to be treated discretely. This is particularly true for the case of transcriptional regulation by transcription factors. In this situation there can be as few as one DNA/transcription factor binding site and mRNA abundances can be as small as 10 or fewer. In contrast, protein abundances can be in the thousands. The technical difficulty with implementing hybrid schemes that include both discrete and continuous random variables is that the Gillespie method requires constant transition rates between reactions. This may not be the case, if some of the chemical species are evolving continuously in time. BioNetS overcomes this problem in one of two ways.

Let $N_d < N$ be the number of discrete chemical species and $M_d \leq M$ the number of reactions that produce a change in one of the N_d chemical species. The overall reaction rate at time t_j for the discrete set of chemical species is

$$\mu_{t_j} = \sum_{i=1}^{M_d} [F_i(t_j) + B_i(t_j)] \tag{27}$$

If the time step Δt for the SDEs is small enough such that

$$p_t = \mu_{t_j} \Delta t < \epsilon << 1 \tag{28}$$

then p_t is approximately the probability of a transition in Δt. In the above equation ϵ is a user-specified tolerance. The probability of two discrete transitions in Δt is proportional to $(\Delta t)^2$. Choosing $\epsilon < 0.1$, which means the probability of two reactions in Δt is less than 0.01, generally produces good results. However, this should be verified on a case-by-case basis. At each time step, BioNetS checks to verify that Ineq. 28 is satisfied for the specified ϵ. If so, a uniform random number R is generated and compared against p_t. If $R < p_t$, then a transition occurred and the conditional probability R/p_t is used to determine which of the discrete transitions occurred. If $p_t > \epsilon$, then the discrete reactions determine the fastest time scale in the system. In this case the Gillespie algorithm is used to update the discrete reactions, and the random time step Δt_j is used to update the SDEs.

3.2.3 Numerical Methods

BioNetS generates code that is tailored to efficiently simulate biochemical reactions. The optimization techniques used by BioNetS allows the software to simulate large systems in reasonable times without requiring high-end computational hardware.

Techniques used to optimize the Gillespie method are:

- For the discrete variables, the program uses data structures that allow only the chemical species and reaction rates that are affected by the current reaction to be updated.

- A bisection search is used to determine which reaction occurred.

The code has both an explicit and a semi-implicit solver, for simulating the chemical Langevin equations. The user specifies at runtime which method to use. By default the semi-implicit solver will be used. The semi-implicit solver uses Newton's method to solve the implicit equations, and for that the program needs to compute the Jacobian and solve a linear system at each iteration. For updating the chemical Langevin equations and hybrid models optimization techniques include:

- The sparse nature of the stoichiometric matrix is used to efficiently store and perform matrix operations.
- After every reaction, only the species and reaction rates affected by that reaction are updated.
- The Jacobian is sparse, and the code takes full advantage of this fact. The program solves and factorizes the Jacobian using sparse methods. Before the code generation, BioNetS computes the entries in the Jacobian symbolically and finds a permutation that decreases the number of fill-ins during the LU factorization. As a result, no zero entries are saved, and the sparse structure is fully exploited. The sparse structure is then used in the LU solve. In the code, no pivots are visible, and no if-statements are left.

3.3 Results and Discussion

In this section, several examples which serve as illustrations of how to use BioNetS and test the accuracy and efficiency of the numerical methods are presented. One particular concern is the accuracy of the Euler methods. While these methods are only of order $\sqrt{\Delta t}$, it is shown that when the approximations that lead to the chemical Langevin equations are valid, the difference between the numerical solutions of the SDEs and the exact discrete Gillespie method are negligible. Currently, the graphical user interface to BioNetS runs on the Macintosh OS X operating system, though the software will generate portable C/C++ code that can be compiled and run in any computing environment. The following examples illustrate the way in which models are entered and run in BioNetS. More detailed documentation is available with the software package.

3.3.1 Dimerization

To begin, consider a simple system that consists of the following two reactions:

$$\emptyset \underset{\delta_m}{\overset{\gamma}{\rightleftharpoons}} \mathcal{M} \tag{29}$$

$$\mathcal{M} + \mathcal{M} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathcal{D} \tag{30}$$

$$\mathcal{D} \xrightarrow{\delta_d} \emptyset \tag{31}$$

In this system, monomer molecules $\mathcal{M}$ are produced at an average rate γ and degraded at an average rate $\delta_m M(t)$. Two monomers can then bind to form a dimer molecule $\mathcal{D}$. The average forward and backward rates for this reaction are $k_1 M(t)(M(t)-1)$ and $k_2 D(t)$, respectively. The dimers are degraded at a rate δ_d. Two cases will be treated. In the first case the cell volume is assumed to be constant, and in the second case the cell is allowed to grow and divide. To model cell growth, the cell volume V_c is treated as a random variable $V_c = \alpha V$, where V is a non-negative discrete random variable and α represents a unit of volume. The random variable V is governed by the reaction

$$\mathcal{V} \overset{k_3}{\rightarrow} \mathcal{V} + \mathcal{V} \tag{32}$$

The above reaction causes V to grow exponentially fast with an average rate of k_3. Note that logistic growth is produced when the backward reaction in Eq. 32 is included.

3.3.2 Constant Volume

Start by considering the simple case in which the volume of the cell remains constant. To use BioNetS follow these steps. Copy BioNetS onto your machine, and double click to launch. Help is included as part of the program, and accessed from the Help menu. The Help document will walk you through all the steps needed to enter reactions and run the simulator.

The user interface asks you to enter the reaction and corresponding rate constants in the top part of the script window. In the bottom part of the script window, you can toggle between panels. The Species panel allows the user to specify how the simulator treats each chemical species, discrete or continuous. The Constants panel lists the order in which the rate constants are referenced. The Output panel allows the user to specify the ouput type. There are two ways to generate program output, either binary or ASCII. Binary output is based on MATLAB binary files, so it is possible to drive the program with MATLAB and use MATLAB's plotting routines to view the output. It is also possible to generate time series and histograms of the data from within BioNetS. Using ASCII files for I/O allows the simulator to be run through shell scripts. The Executable panel allows the user to generate either an executable file or source code. BioNetS generates portable C/C++ code that can be compiled and run in any computing environment. BioNetS can directly compile the C/C++ code. However, this requires the Developer tools, included on all recent Apple machines and available directly from http://developer.apple.com for free. The compiled code can then be run from within BioNetS. The Comments panel is available for the user to enter descriptive comments about the model.

To run BioNetS as a BioSpice agent, you need to move the source directory onto a OAA-supported system. Once there, open up the MakeOAA file and specify the locations of your oaalib folder. Then just type “make -f MakeOAA” (without the quotes) to create the agent.

Simluations indicated that the agreement between the two different methods is very good. These findings indicate that the chemical Langevin equations are accurately capturing the dynamics and steady-state behavior of the discrete system.

3.3.3 Cell Growth and Division

In this section, it is described how cell growth and division can be modeled using BioNetS. It is assumed that the cell is experiencing exponential growth up until the time it divides. As discussed above, the cell volume V_c is treated as a random variable. In this model cell division occurs when V_c exceeds a threshold value V_{max}. When cell division occurs the volume is halved, and the proteins are randomly divided between the two cells using a binomial distribution. Only one of the daughter cells is tracked. Because second-order reactions require two molecules to collide, the rate constants for these reactions should scale like $k_1 = k_1'/V_c$. It is also assumed that the production rate of monomers scales as $\gamma = \gamma' V_c$. This is a reasonable assumption, because as the cell grows the transcription and translation machinery increases. These assumptions produce the following rate equations for the concentrations

$$\begin{aligned}
\frac{d[M]}{dt} &= -2k_1'[M]^2 + 2k_2[D] + \gamma' - (\delta_m + k_3)[M] && (33)\\
\frac{d[D]}{dt} &= k_1'[M]^2 - k_2[D] - (k_3 + \delta_d)[D] && (34)\\
\frac{dV_c}{dt} &= k_3 V_c && (35)
\end{aligned}$$

The terms in Eqs. 33 and 34 that involve k_3 arise because of dilution due to cell growth. The same parameter values as in the constant volume case are used except $\delta_m = 1$ and $\delta_d = 0$. The cell growth rate $k_3 = 0.02$ and the scale factor for the volume, α, is equal to 1. With these choices of parameter values, we expect the average behavior of this system to be similar to that of the constant volume case. Simulations for this simple example indicated that the main effect of volume growth is to act as an additional noise source and increase the variability of the distributions.

3.3.4 A Chemical Oscillator

BioNetS is next used to simulate a two-gene system that has been studied in the literature. In this system, the protein $\mathcal{A}$ coded for by gene a acts as an activator for gene a and gene r, by binding to the promoter regions, $\mathcal{P}_a$ and $\mathcal{P}_r$, of the respective gene. This increases the rate of $m\mathcal{RNA}_a$ and $m\mathcal{RNA}_r$ production by a factor α_a and α_r, respectively. The protein $\mathcal{R}$ acts as a repressor for both genes by binding to $\mathcal{A}$ to form the inactive complex $\mathcal{A}\_\mathcal{R}$. All gene products, mRNA and protein, are actively degraded. However, the heterodimer $\mathcal{A}\_\mathcal{R}$ protects the $\mathcal{R}$ subunit from degradation. The system consists of 9 chemical species and the following 14 biochemical reactions:

$$\mathcal{P}_a \xrightarrow{k_1} \mathcal{P}_a + m\mathcal{RNA}_a \tag{36}$$

$$\mathcal{P}_a\_\mathcal{A} \xrightarrow{\alpha_a k_1} \mathcal{P}_a\_\mathcal{A} + m\mathcal{RNA}_a \tag{37}$$

$$\mathcal{P}_r \xrightarrow{k_2} \mathcal{P}_r + m\mathcal{RNA}_r \tag{38}$$

$$\mathcal{P}_r\_\mathcal{A} \xrightarrow{\alpha_r k_2} \mathcal{P}_r\_\mathcal{A} + m\mathcal{RNA}_r \tag{39}$$

$$m\mathcal{RNA}_a \xrightarrow{k_3} m\mathcal{RNA}_a + \mathcal{A} \tag{40}$$

$$m\mathcal{RNA}_r \xrightarrow{k_4} m\mathcal{RNA}_r + \mathcal{R} \tag{41}$$

$$\mathcal{A} + \mathcal{R} \underset{k_6}{\overset{k_5}{\rightleftharpoons}} \mathcal{A}\_\mathcal{R} \tag{42}$$

$$\mathcal{P}_a + \mathcal{A} \underset{k_8}{\overset{k_7}{\rightleftharpoons}} \mathcal{P}_a\_\mathcal{A} \tag{43}$$

$$\mathcal{P}_r + \mathcal{A} \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} \mathcal{P}_r\_\mathcal{A} \tag{44}$$

$$\mathcal{A} \xrightarrow{k_{11}} \emptyset \tag{45}$$

$$\mathcal{R} \xrightarrow{k_{12}} \emptyset \tag{46}$$

$$m\mathcal{RNA}_a \xrightarrow{k_{13}} \emptyset \tag{47}$$

$$m\mathcal{RNA}_r \xrightarrow{k_{14}} \emptyset \tag{48}$$

$$\mathcal{A}\_\mathcal{R} \xrightarrow{k_{15}} \mathcal{R} \tag{49}$$

An interesting feature of the system is that it is capable of producing sustained oscillations.

The chemical species $\mathcal{P}_a$, $\mathcal{P}_r$, $\mathcal{P}_r\_\mathcal{A}$, and $\mathcal{P}_r\_A$ are binary random variables: they can only take on the values 0 or 1. Therefore, these species cannot be approximated as continuous random variables. All the other chemical species appear in sufficient quantities to justify the continuum approximation. The hybrid model was run using the semi-implicit Euler method, and for these parameter values, runs three times faster than full model. Visually, the agreement between the two methods appears good. To test the accuracy of the Euler method, BioNetS was used to construct 2-D histograms of R versus $mRNA_r$.

Simulations showed excellent agreement between the discrete and hybrid models. This indicates that the hybrid model is accurately sampling the steady-state distribution. To verify that the hybrid model faithfully captures the dynamics of the system, the power spectra of both models were computed. Again, excellent agreement was found between the discrete and hybrid models.

3.3.5 An Engineered Promoter System

Genetic regulatory networks consist of sets of genes whose levels of expression in a cell are interdependent. This dependence arises through the action of transcription factors, proteins which bind to operator sites on the DNA strand and influence the rate at which a gene product is generated. Once bound, these regulatory proteins affect the binding affinity of RNA polymerase, an enzyme that binds to promoter sequences in the DNA and initiates transcription of messenger RNA (mRNA) strands, which are subsequently translated into proteins. These proteins may themselves act as transcription factors, influencing their own rates of expression or those of other gene products and thus forming networks of connected genes. Using standard techniques in modern molecular biology, it is possible to design novel systems of promoter-gene pairs, such that virtually any desired regulatory network architecture may be instantiated; such networks are often called "synthetic gene networks." Recent implementations have included direct negative and positive feedback, a bistable switch, an oscillator, an intercellular communication system, and a bimodal self-activating system.

In this example, BioNetS is used to implement a model of a simple, open-loop network based around a novel engineered promoter, which was designed and constructed by the project team. The promoter, called O_RO_{lac}, combines the O_{lac}, O_R1, and O_R2 operator sites, so that it is repressed by the lac repressor protein (LacI) and activated by the lambda repressor protein (CI). Experiments were conducted in which the promoter, along with other sites to produce the activator and repressor proteins, is integrated into a high copy number plasmid and inserted into a strain of *Escherichia coli*. The promoter's activity is observed using a fluorescent reporter, Green Fluorescent Protein (GFP). The goal here is to provide a reasonably complex test case to evaluate the performance of BioNetS.

The processes to be captured by the model are: transcription and degradation of mRNA strands; translation of mRNA into protein; degradation of protein; formation of protein multimers (dimers in the case of CI, tetramers in the case of LacI); LacI binding to isopropyl-β-D-thiogalactopyranoside (IPTG), a chemical inducer that reduces LacI's binding affinity for O_{lac}; and protein-DNA binding at the O_RO_{lac} promoter's operator sites. We define the following chemical species: G, GFP; M_g, mRNA coding for GFP; X, CI monomer; X_2, CI dimer; M_x, mRNA coding for CI; D_x, the arabinose-inducible $pBAD$ promoter site producing CI; Y, LacI monomer; Y_2, LacI dimer; Y_4, LacI tetramer; I_0, IPTG (present in massive excess and thus taken to be constant); Y_I, LacI tetramer bound to IPTG; M_y, mRNA coding for LacI; and D_y, the P_LtetO1 site constitutively producing LacI. In addition to these, species D_0 through D_8 are defined, representing the various permutations of proteins bound to the three operator sites in the O_RO_{lac} promoter. There are twelve combinatorial possibilities, but three of them are eliminated on the basis that CI (X_2) binding O$_R$2 but not O$_R$1 is unlikely, because of the low binding affinitity of CI for O$_R$2 compared to O$_R$1. This reflects the regulatory effect of the proteins; for example, CI bound to O_R2 leads to a 10-fold increase in transcription rate, while LacI bound to O_{lac} halts transcription completely (note that we assume in the event of simultaneous binding of activator and repressor, repression "wins" and transcription is halted).

The following irreversible reactions represent the processes of transcription, translation, and degradation:

$$\mathcal{D}_0 \overset{\beta_g}{\to} \mathcal{D}_0 + \mathcal{M}_g \tag{50}$$
$$\mathcal{D}_1 \overset{\beta_g}{\to} \mathcal{D}_1 + \mathcal{M}_g \tag{51}$$
$$\mathcal{D}_2 \overset{10\beta_g}{\to} \mathcal{D}_2 + \mathcal{M}_g \tag{52}$$
$$\mathcal{D}_y \overset{\beta_y}{\to} \mathcal{D}_y + \mathcal{M}_y \tag{53}$$
$$\mathcal{D}_x \overset{\beta_x}{\to} \mathcal{D}_x + \mathcal{M}_x \tag{54}$$
$$\mathcal{M}_g \overset{\beta_T}{\to} \mathcal{M}_g + \mathcal{G} \tag{55}$$
$$\mathcal{M}_y \overset{\beta_T}{\to} \mathcal{M}_y + \mathcal{Y} \tag{56}$$
$$\mathcal{M}_x \overset{\beta_T}{\to} \mathcal{M}_x + \mathcal{X} \tag{57}$$
$$\mathcal{M}_g \overset{\gamma_{mrna}}{\to} \emptyset \tag{58}$$
$$\mathcal{M}_y \overset{\gamma_{mrna}}{\to} \emptyset \tag{59}$$
$$\mathcal{M}_x \overset{\gamma_{mrna}}{\to} \emptyset \tag{60}$$
$$\mathcal{G} \overset{\gamma_{prot}}{\to} \emptyset \tag{61}$$
$$\mathcal{Y} \overset{\gamma_{prot}}{\to} \emptyset \tag{62}$$
$$\mathcal{X} \overset{\gamma_{prot}}{\to} \emptyset \tag{63}$$

As in previous reactions, the caligraphic letters represent individual molecules of each species. All times and rates are scaled by the cell division time.

Experimental measurements generally provide equilibrium rather than rate constants, and thus when writing reversible reactions we use the following notational convention: a reaction with equilibrium constant K has forward rate constant KR and backward rate constant R, where R is a scaling factor which sets the speed at which the reaction approaches equilibrium (three values of R – 1, 10, and 100 – are considered). Using this notation, protein-protein binding is represented with the following set of reactions

$$\mathcal{Y} + \mathcal{Y} \overset{K_y}{\rightleftharpoons} \mathcal{Y}_2 \tag{64}$$

$$\mathcal{Y}_2 + \mathcal{Y}_2 \overset{K_{y2}}{\rightleftharpoons} \mathcal{Y}_4 \tag{65}$$

$$\mathcal{Y}_4 + I_0 \overset{K_{yI}}{\rightleftharpoons} \mathcal{Y}_I \tag{66}$$

$$\mathcal{X} + \mathcal{X} \overset{K_x}{\rightleftharpoons} \mathcal{X}_2 \tag{67}$$

Finally, protein-DNA binding is given by:

$$\mathcal{D}_0 + \mathcal{X}_2 \overset{K_1}{\rightleftharpoons} \mathcal{D}_1 \tag{68}$$

$$\mathcal{D}_1 + \mathcal{X}_2 \overset{K_2}{\rightleftharpoons} \mathcal{D}_2 \tag{69}$$

$$\mathcal{D}_0 + \mathcal{Y}_4 \overset{K_3}{\rightleftharpoons} \mathcal{D}_3 \tag{70}$$

$$\mathcal{D}_1 + \mathcal{Y}_4 \overset{K_3}{\rightleftharpoons} \mathcal{D}_4 \tag{71}$$

$$\mathcal{D}_3 + \mathcal{X}_2 \overset{K_1}{\rightleftharpoons} \mathcal{D}_4 \tag{72}$$

$$\mathcal{D}_2 + \mathcal{Y}_4 \overset{K_3}{\rightleftharpoons} \mathcal{D}_5 \tag{73}$$

$$\mathcal{D}_4 + \mathcal{X}_2 \overset{K_2}{\rightleftharpoons} \mathcal{D}_5 \tag{74}$$

$$\mathcal{D}_0 + \mathcal{Y}_I \overset{K_4}{\rightleftharpoons} \mathcal{D}_6 \tag{75}$$

$$\mathcal{D}_1 + \mathcal{Y}_I \overset{K_4}{\rightleftharpoons} \mathcal{D}_7 \tag{76}$$

$$\mathcal{D}_6 + \mathcal{X}_2 \overset{K_1}{\rightleftharpoons} \mathcal{D}_7 \tag{77}$$

$$\mathcal{D}_2 + \mathcal{Y}_I \overset{K_4}{\rightleftharpoons} \mathcal{D}_8 \tag{78}$$

$$\mathcal{D}_7 + \mathcal{X}_2 \overset{K_2}{\rightleftharpoons} \mathcal{D}_8 \tag{79}$$

In all, the system consists of 21 species, participating in 34 reactions. The reactions are entered into BioNetS using the same method described in the previous examples. BioNetS' ability to represent individual species as either discrete or continuous is used to formulate three versions of the model: fully discrete, fully continuous, and a hybrid version in which the DNA species D_0 through D_8 are discrete while all other species are continuous. The value of R, the scaling factor for reversible reactions is varied, and all other parameters are kept fixed at the following nondimensionalized values: $\beta_g = 0.1$, $\beta_y = 1$, $\beta_x = 0.5$, $\beta_T = 10$, $\gamma_{mrna} = 3.5$, $\gamma_{prot} = 0.7$, $K_y = 0.01$, $K_{y2} = 0.1$, $K_{yI} = 2 \times 10^{-6}$, $K_x = 0.05$, $K_1 = 0.3$, $K_2 = 2K_1$, $K_3 = 0.008$, $K_4 = 1.4 \times 10^{-4} K_3$, $I_0 = 1 \times 10^6$.

To evaluate the steady-state probability distributions produced by the reaction system, simulations 250000 cell cycles in length were used to accumulate histograms (a built-in feature of BioNetS) of the number of molecules of GFP (species G), for each of the three versions of the model. The resulting distributions were essentially identical, indicating that the continuum approximations used in the fully continuous and hybrid forms of the model were valid. Not all species in the system are well approximated as continuous variables The fully continuous model fluctuates into negative values, indicating that the continuum approximation has broken down. This does not significantly affect the distribution for GFP because the other, more common DNA states dominate the system's behavior; note, however, that if genomic DNA were considered rather than a high copy number plasmid, one would not be able to employ a fully continuous model. The hybrid model, by treating the DNA species as continuous, eliminates the fluctuations into negative values. In general, the appropriate approximations will depend on both the system and the variables of interest: in the present example, if one were interested in the behavior of the operator sites themselves, one would not be able to use the fully continuous version of the model, but as a model solely of GFP expression the approximation suffices. Comparisons between types of models should be made to test the underlying assumptions, and BioNetS facilitates this process.

Simulations 200 cell cycles in length were used to test the speed at which the three model versions ran. In each case, 200 simulations were run using a consistent set of 200 different random seeds; all runs were started with identical initial conditions. For the fully continuous and hybrid systems, the semi-implicit scheme was numerically stable and yielded consistent histograms for all time step sizes between $dt = 0.001$ and $dt = 0.5$, but the latter corresponds to just two time points per cell division cycle (recall that all times are scaled by the cell division time), and it was chosen instead to sample 20 points per cycle and set $dt = 0.05$. Simulations showed that the fully continuous method was always fastest, with the degree of improvement over the exact, fully discrete method depending strongly on the value of R, the scaling factor for the reversible reaction rates. For $R = 1$, the fully continuous method was only 1.4-fold faster than the fully discrete method, but as R is increased this speed advantage increases to over 4-fold at $R = 10$, then to over 30-fold at $R = 100$. (Note that the speed advantage of the fully continuous over the fully discrete method increases with the abundances of the chemical species. Shifting parameters to generate higher protein numbers can yield cases in which the continuum approximation is hundreds of times faster than the discrete approach; runs not shown here.) Use of a hybrid discrete/continuous method did not, for this particular model system, offer any speed gain over the fully discrete approach; the increased time involved in computing the Jacobian for the semi-implicit method is more time-consuming than simply simulating the reactions directly. Optimizing efficiency requires testing various potential approaches, and BioNetS makes this a simple process.

3.4 Conclusions

BioNetS was developed to be a reliable tool for studying the stochastic dynamics of large chemical networks. The software allows the user to specify which of the chemical species in the network should be treated as discrete random variables and which can be approximated as continuous random variables. The software is highly optimized for speed and should be be able to simulate networks consisting of hundreds of chemical species. The accuracy of the numerical methods was verified by considering several test systems (a dimerization reaction, a chemical oscillator, and an engineered promoter), each of which shows excellent agreement between the fully discrete version and the fully or partially continuous versions. BioNetS, by providing a simple, user-friendly interface, will allow biological experimentalists to formulate biochemical reaction models of their systems quickly and easily, ideally increasing the number of systems in which direct comparisons are available between models and experimental results. Clearly, not every possible biological system can be captured in the current version of BioNetS, and its capabilities will continue to grow in the future.

3.5 Requirements

- **Project name**: BIOchemical NETwork Stochastic Simulator (BioNetS)
- **Operating system**:
  - User interface: Macintosh OS X, version 10.2 or above.
  - Generated source code: Ability to compile portable C++ code. Makefiles included for OS X and Linux.
- **Programming language**: C++.
- **Other requirements**: None.

4 Project Publications

- Gardner TS, di Bernardo D, Lorenz D and Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301: 102–105 (2003).
- Kaern M, Blake WJ and Collins JJ. The engineering of gene regulatory networks. *Annual Reviews of Biomedical Engineering* 5: 179–206 (2003).
- Prediction and measurement of an autoregulatory genetic module. F. Isaacs, J. Hasty, C. Cantor and J.J. Collins. *Proc. Natl Acad. Sci. USA* 100: 7714–7719 (2003).

- Reverse engineering gene networks — integrating genetic perturbations with dynamical modeling. J. Tegner, M.K.S. Yeung, J. Hasty and J.J. Collins. *Proc. Natl Acad. Sci. USA* 100: 5944–5949 (2003).

- Noise in eukaryotic gene expression. W. Blake, M. Kaern, C. Cantor and J.J. Collins. *Nature* 422: 633–637 (2003).

- Engineered gene circuits. J. Hasty, D. McMillen, and J.J. Collins. *Nature* 420: 224–230 (2002).

- Synchronizing genetic relaxation oscillators with intercell signaling. D. McMillen, N. Kopell, J. Hasty, and J.J. Collins. *Proc. Natl Acad. Sci. USA* 99: 679–684 (2002).

- Reverse engineering gene networks using singular value decomposition and robust regression. M.K.S. Yeung, J. Tegner, and J.J. Collins. *Proc. Natl Acad. Sci. USA* 99: 6163–6168 (2002).

- Translating the noise. J. Hasty and J.J. Collins. *Nature Genetics* 31: 13–14 (2002).

- Intrinsic noise in gene regulatory networks. M. Thattai and A. van Oudenaarden. *Proc. Natl Acad. Sci. USA* 98: 8614-8619 (2002).

- Regulation of noise in the expression of a single gene. E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. *Nature Genetics* 31: 69-73 (2002).

- Attenuation of noise in ultrasensitive signalling cascades. M. Thattai and A. van Oudenaarden. *Biophysical Journal* 82: 2943-2950 (2002).

- Frequent Sampling Reveals Dynamic Responses by the Transcriptome to Routine Media Replacement in HepG2 Cells. K. Morgan, W. Casey, M. Easton, D. Creech, H. Ni, L. Yoon, S. Anderson, C. Qualls, L. Crosby, P. Bloomfield, A. MacPherson, and T. Elston. *Tox. Path.*, 32:448-461, 2003.

- A robust numerical algorithm for studying biomolecular transport processes. H. Wang, C. Peskin and T. Elston. *J. Theor. Biol.*, 221:491-511, 2003.